

Social Security Administration's Data Inventory and Plan for Releasing High Value Data

Our approach for supporting data
transparency and accountability

September 30, 2010



Contents

Table of Contents

Introduction.....	3
New Open Government Requirements.....	3
Actions Taken.....	4
List of Datasets in Open Government Plan	5
Background on Enterprise Data and Our Business Intelligence Architecture	9
Methodology for Obtaining Datasets for Data.gov	10
Additional Datasets for Data.gov	11
Additional SSA Data-Fostering the Use of Data and Other Traditional Releases	13
Additional Future Sources and Types of Information Which May be High-Value	14
Concluding Thoughts	15
Appendix -Checklist for Data.gov Submission	16
Charts	
Chart 1 --Social Security Datasets on Data.gov.....	6
Chart 2 -- Social Security's New Planned High-Value Datasets and Information for Data.gov (For posting in 2010 and 2011).....	8
Chart 3 -- Examples of Open Government High-Value Dataset Development	10
Chart 4 -- Future New High-Value Datasets (Release Begins in 2012).....	11
Chart 5 -- Information That May Be High-Value but Not Currently Available (Planning for Release in 2013 and Continuing	14

Social Security Administration's Data Inventory and Plan for Releasing High-Value Data September 2010

“Transparency promotes accountability and provides information for citizens about what their Government is doing.”

--Barack Obama

Introduction

Social Security celebrated its 75th birthday on August 14, 2010. Over the last 75 years, we have collected data to carry out our mission. Our data are about people-their wages, their identifying information, their employers, their addresses, and much more. The first regulation we published included a commitment to the public to safeguard the personal information entrusted to us. This commitment is as solid as it was 75 years ago and is further strengthened by privacy laws. We cannot publicly release much of our data because it is protected by the Privacy Act, the Internal Revenue code, the HIPAA, and other statutes. While some of the data can be anonymized, much of it cannot. Our inventory planning recognizes these constraints and all releases will protect privacy in accordance with all applicable laws.

New Open Government Requirements

In the first days of his administration, President Obama issued a memorandum for the heads of executive departments and agencies announcing a commitment to transparency and open government. In this document, the President instructed the Director of the Office of Management and Budget (OMB) to issue instructions for implementing the open government principles. The Open Government Directive (M-10-06) released on December 8, 2009 provided specific requirements agencies and departments must follow in order to fully adopt the open government principles. It also mandated the development and publication of an Open Government Strategic Plan.

Under the transparency principle, OMB directed agencies to explain in detail in their Open Government Plan how they would improve transparency. Specifically, the Directive said, “To increase accountability, promote informed participation by the public, and create economic opportunity, each agency shall take prompt steps to expand access to information by making it available online in open formats.” We recognize that the foundational steps to becoming more transparent and increasing accountability are to inventory agency high-value¹ information currently available for download, to identify high-value information not yet available; and establish a reasonable timeline for publication online in open formats.

¹ High value information is information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.

Actions Taken

We have completed a number of actions to increase transparency by making information available at Socialsecurity.gov/open and Data.gov.

- In response to the launch of Data.gov in fiscal year (FY) 2009, we began our inventory of data and released two datasets in May 2009. (See last two items in Chart 1.) These datasets met government standards for Data.gov as well as our privacy and disclosure review processes.
- During 2009, we continued our inventory of data. We addressed disclosure and privacy issues, in part, by establishing a point of contact in the Office of the Chief Information Officer for review of each dataset prior to release to Data.gov. We established an executive steering committee for oversight of all open government activities.
- In January 2010, as a result of our ongoing inventory process, we released additional, high-value datasets in accordance with the Open Government Directive.
- On February 6, 2010, we launched our new Open Government Webpage (Socialsecurity.gov/open), containing datasets and information about our overall management and organizational structure. Through this webpage, we also launched the IdeaScale tool that we used to obtain public input on our planned transparency and data initiatives. After a period of public engagement (that included the IdeaScale tool feedback, advocate meetings, and researchers' input), our executive steering committee evaluated the input and recommended goals, initiatives, objectives, and milestones.
- On June 24, 2010, we published our Open Government Plan which included a list of datasets already posted or planned through FY 2011. In this document, Chart 1 contains datasets already posted and Chart 2 lists planned dataset releases through 2011.
- The Open Government Plan includes five major objectives that support the goal of increasing transparency. One of the critical activities is a data inventory, which was completed in September 2010, the milestone shown for this activity in the Open Government Plan.
- In response to the Open Government Directive, we named an Executive Accountable for Publicly Disseminated Federal Spending Information Integrity. We also released our Data Quality Plan for Federal Spending Information on May 14, 2010. The Data Quality Plan outlines our strategy beginning with the certification of our financial data, extending to our other data as well. The long-term strategy consists of a data quality framework that includes:
 - governance structures (e.g., disclosure review boards, executive steering committee, working groups, etc.) and processes;
 - risk assessments;
 - control activities;
 - communications; and
 - monitoring.
- We established a working group of subject matter experts that meets to review release timeframes, quality, security, and privacy. Following the leadership of the Program Management Office of Data.gov at the General Services Administration (GSA), we developed a dataset checklist sheet modeled after the one GSA uses (see Appendix). Included in the checklist is the foundational element of quality as required by the Information Quality Act:
<http://www.fws.gov/informationquality/section515.html>.

- We issued implementation instructions that address data utility, objectivity, integrity, transparency and reproducibility, and made this publicly available at:

www.ssa.gov/515/ssaguidelines.html/

- We examined our disclosure review board process to ensure it conforms to Data.gov standards. We connected the review board management and our Chief Privacy Officer with the OMB and White House Working Group on Privacy and Security. They ensure that we are aware of national security and privacy issues in the Federal community, enabling us to take appropriate action.
- We refreshed our understanding of the legal limitations of sharing certain data such as master earnings file data protected under the Internal Revenue Code.

List of Datasets in Open Government Plan

We posted [22 datasets on Data.gov](#). All datasets relate to our core mission and align with our strategic goals and performance measures. The following information in Charts 1 and 2 shows the datasets already posted as well as those scheduled for release by the end of FY 2011.

Charts

Chart 1--Social Security Datasets on Data.gov

Datasets	How Social Security Meets High-Value Criteria				
	Increase agency accountability and responsiveness	Improve public knowledge of agency and its operations	Further core mission of agency	Create economic opportunity	Respond to identified public need and demand
Hearing Process <ul style="list-style-type: none"> • Administrative Law Judge (ALJ) dispositions • Hearing Office dispositions • Hearings held in person or via Video Conferencing • Hearing Office Average Processing Time • Hearing Office Workload Data • NetStat Report (average time from the hearing request date until a hearing is held) 	◆	◆	◆		◆
Disability Decisions <ul style="list-style-type: none"> • State Disability Determination Services Workload Data • Social Security Disability Data 	◆	◆	◆		◆
2008 Freedom of Information Act (FOIA) Annual Report 2009 FOIA Annual Report	◆				◆
National Beneficiary Survey <ul style="list-style-type: none"> • Round 1 • Round 2 • Round 3 	◆	◆	◆		◆
Average Monthly Payments for Supplemental Security Income , by State or Other Area, Eligibility Category, and Age, December 2008					◆
Supplemental Security Income Payments by Type of Payment, Sex, Eligibility Category, and Age, December 2008					◆
Supplemental Security Income Payments, Recipients by State and Other Area, Eligibility Category, and Age, December 2008					◆
Average Monthly Social Security Benefits of Disabled Beneficiaries and Nondisabled Dependents by Basis of Entitlement, Age, and Sex, December 2008					◆

Number of Disabled Workers Receiving Social Security Benefits by Sex, State or Other Area, and Age, December 2008					◆
Supplemental Security Income Public-Use Microdata File, 2001 Data	◆	◆			
Old-Age, Survivors, and Disability Insurance Public-Use Microdata File, 2001 Data	◆	◆			
Benefits Data from the Benefits and Earnings Public-Use file, 2004 <i>(Social Security Benefit Information of Beneficiaries only)</i>	◆	◆			
Earnings Data from the benefits and Earnings Public-Use File, 2004 <i>(Longitudinal Earnings Information of Beneficiaries only)</i>	◆	◆			

**Chart 2--Social Security's New Planned High-Value Datasets and Information for Data.gov
(For posting in 2010 and 2011)**

High-Value Criteria					
	Increase agency accountability and responsiveness	Improve public knowledge of agency and its operations	Further core mission of agency	Create economic opportunity	Respond to identified public need and demand
New Disability Determination Services Processing <ul style="list-style-type: none"> • Disability Reconsideration Data for: <ul style="list-style-type: none"> ○ Social Security Claims ○ Supplemental Security Income Claims with breakout for disabled children claims ○ Simultaneously filed or Concurrent Claims • Continuing Disability Review Data for: <ul style="list-style-type: none"> ○ Social Security Claims ○ Supplemental Security Income Claims ○ Simultaneously filed or Concurrent claims • Prototype Case Workload by Participating State Agency (<i>cases in which reconsideration was not held</i>) • Disability Workload Processed by Federal Entity State Disability Determination Services Budget Information State Disability Determination Services Processing Time Number and Percentage of Quick Disability Allowances Number and Percentage of Compassionate Allowances	◆	◆	◆		◆
Freedom of Information Act Report for 2010	◆				◆
Retirement Claims Filed and Cleared (aggregated information)	◆	◆	◆		◆
Number and Percentage of Retirement Claims Filed via Internet	◆	◆	◆		◆
Internet Usage for Selected Online Transactions	◆	◆	◆		◆
Quality Workload Statistics	◆	◆	◆		◆
Earnings Public Use File (<i>Demographic and earnings information for a sample of all social security numbers</i>)	◆	◆			
Benefits and Earnings Public Use File updated (<i>Social Security Benefit Information and Longitudinal Earnings Information of Beneficiaries only</i>)	◆	◆			
National Survey of Children and Families	◆	◆			

Datasets from Statistical Modernization Initiative	◆	◆			
Field Office Waiting Time	◆	◆	◆		◆
Social Security 800 Number Call Volume and Busy Rate	◆	◆	◆		◆
Speed in Answering Social Security 800 Number Calls	◆	◆	◆		◆

Background on Enterprise Data and Our Business Intelligence Architecture that Impacts Data for Data.gov

As our program responsibilities grow, data and databases also grow and change. Historically, we developed our programmatic software at different times for different purposes, and deployed it in a stovepipe environment.² In order to give a comprehensive and transparent picture of our data, we will need to derive datasets for Data.gov from several sources (e.g. Case Processing Management System, Appeals Review Processing System, etc.) as we continue conversion from our older, homegrown database structures {e.g. Master Data Access Method (MADAM)} to modern relational database technologies.

Going forward we will explore incorporating data transparency considerations into the systems development lifecycle so that providing high-value data in standard formats for public use is a more efficient, automated process. Building steps into the lifecycle involves working with and through several governance groups. For example, the lifecycle Change Control Boards determine the appropriate placement of each step in the lifecycle relative to other tasks and data needs; the Deputy Commissioner for Systems Management Steering Committee determines the resource and schedule impact on project activities. We tentatively plan to identify a selection of pilot projects and evaluate their results prior to implementing any significant lifecycle changes.

In addition to the database strategy referenced in the paragraph above, we have made progress in overhauling information-sharing through continued development of a business intelligence (BI) architecture. The process began in the late 1990's with internal collaboration and the use of technology solutions to address data integration and workload management challenges. Once the BI architecture was in place, we transitioned workloads incrementally, workload by workload, based upon value to us, and governed by our architectural review board. Master data management is an important component of building the BI architecture and provides consistency and data quality.³ It is through the BI architecture that we will be able to produce summarized datasets for Data.gov.

Methodology for Obtaining Datasets for Data.gov

Chart 3 shows three different methods for deriving datasets for Data.gov. Line 1 reflects the use of the BI architecture. Using master data management in the BI architecture, we summarize and aggregate data, making it available for Data.gov without personally identifiable information (PII). Besides the datasets available through the BI architecture model, we produce additional datasets through data extraction from the program

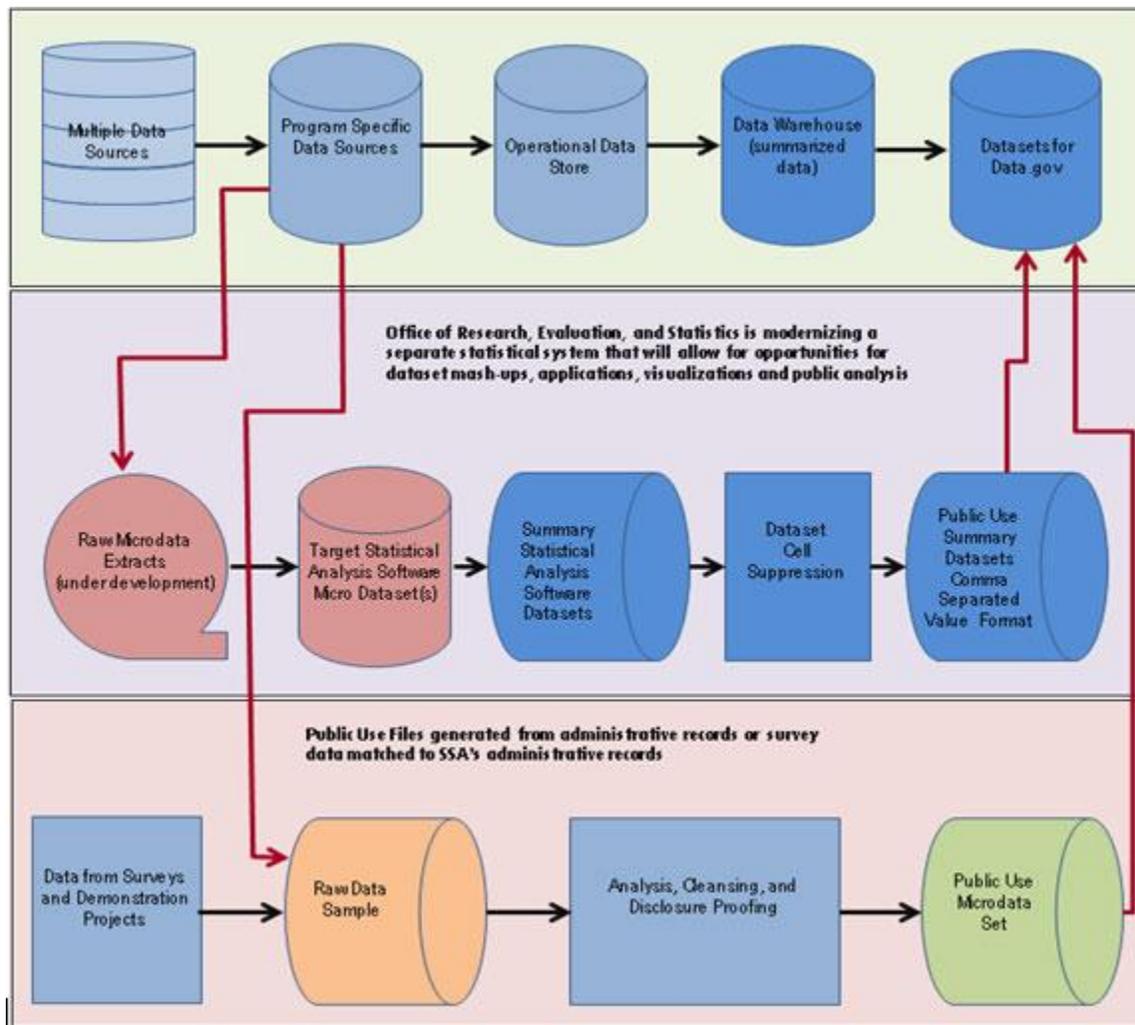
² Social Security Administration, Information Technology Vision 2009-2014, February 2009, page 14.

³ Case Study: BI Strategy Prepares US Social Security Administration for the Future, Gartner Research, Bill Gassman, July 31, 2009, page 4.

specific data we maintain. Line two shows how we use anonymization methods to develop public use files as well as summarized reports. Line three illustrates a similar process to produce public reports from the survey data we collect, along with program specific data we maintain.

While not reflected in the Chart 3 diagram, we also produce datasets in response to Freedom of Information Act (FOIA) requests. Since we generate the data and information in response to individual requests, there is not just one method for responding to FOIA requests. We look at each information release to determine if it is appropriate for Data.gov.

Chart 3—Examples of Open Government High-Value Dataset Development⁴



Additional Datasets for Data.gov Based Upon Inventory of Existing High-Value Data

Based upon a review of our data, we identified high-value data categories, shown below, that will improve public knowledge of our operations and our programs. For all of the categories, we will extract reports that

⁴ This chart does not include datasets produced in response to FOIA requests. The numerous datasets related to our budget process, Annual Performance Plan, Performance.Gov and Priority Goals are also not included.

contain summarized data so that even if the public combines our datasets with other data, no violations of privacy can occur. We will verify and certify the quality of the data in accordance with the Information Quality Act and the Data Quality Framework developed by our Executive Accountable for Publicly Disseminated Federal Spending Information Integrity. In addition, our Open Government Executive Steering Committee will advise, and as needed, concur, on the appropriate level of data (local, regional, national, etc.) and frequency of release (e.g. weekly, annually, etc). As needed, we will refer datasets to disclosure review boards for clearance should there be any questions regarding security and/or privacy. The Checklist for Data.gov Submission (see Appendix) will be signed off by the appropriate officials before the data are made available to Data.gov.

The following list is grouped according to major mission and program areas, not by priority. These areas have high-value data that are available and not released. These areas supplement the extensive data already released on Disability Determination Services and Office of Disability Adjudication and Review case processing, as well as statistical information that our researchers make available on a regular basis. Our first priority for earliest release, in FY 2012, will be the data that have already been determined to be the “official agency measure” in our Unified Measurement System and available through the BI Architecture, as well as data in support of [Performance.gov](#) metrics. The second priority will be budget information, in consultation with the Senior Accountable Official on the Quality of Federal Spending Information. The third priority will be administrative information that includes data about staffing and human resources. By focusing efforts in these areas, by the end of FY 2012, we will provide a transparent picture of program activities, our budget, and our staff. We will also release as soon as possible, without respect to the above priorities, those datasets requested by our key audiences. (See our Communications Plan, Appendix C, in the Open Government Plan.)

The following chart reflects the inventory of high-value data not released, but available as explained in the paragraph above.

Chart 4--Future New High-Value Datasets (Release Begins in 2012)

Service

Field office visitors

Field office claims appointments scheduled within 21 days

Public satisfaction with field office and 800 number service

Initial Claims

Field office claims pending

Number and percentage of disability claims filed via the Internet

Disability Determination Services production per workyear

Disability Determination Services case processing time with health information technology

Old-Age, Survivors, and Disability Insurance and Supplemental Security Income case accuracy and dollar accuracy

Disability Determination Services net accuracy

Hearings

Appeals Council and Court remands sent to hearing level

Hearing production per workyear total and decision writing

Hearings pending by electronic or paper folders

Appeals Council

Appeals Council request for review receipts, dispositions, pending

Appeals Council production per workyear

Court remands sent to Appeals Council

Court Cases

Court level receipts, completed and pending

Earnings

Annual earnings items completed by electronic and paper

Forms W-2 completed (all formats)

Other Workloads

Continuing disability review case outcomes, completed and pending

Supplemental Security Income redeterminations processed compared to budgeted target for the year

Enumerations completed, processing time, and enumeration at birth completed

Medicare Part B and Part D workloads

Claimant representative data

Old-Age, Survivors, and Disability Insurance notice data

Limited English Proficiency data

Administrative

Social Security staff on duty with impact to national and state economy

Disability Determination Services staff on duty with impact to national and state economy

Other

Performance.gov metrics data

Administrative cost data by appropriation and allotment, which may include the following (Disability Determination Services budget data already provided for in Open Government Plan Appendix A):

Information Technology Systems	Social Security Advisory Board
Reimbursable activity	Automation Investment Fund
Delegated Buildings	Disaster Relief
Construction	Recovery Act funds
Low Income Subsidy	Office of the Inspector General
Continuing disability review allotments	Other allotments and targeted appropriations
Supplemental Security Income non-disability redetermination allotments	

Additional SSA Data-Fostering the Use of Data and Other Traditional Releases

1. The Research and Statistical Community

Continuing our long tradition of work in the research community, we will develop and release additional statistical tables, studies and research working papers as part of our ongoing mission. Research and statistical products by our Office of Research, Evaluation, and Statistics are available at: <http://www.ssa.gov/policy/>. Our regular publications include, for example, the *Annual Statistical Supplement to the Social Security Bulletin*. The *Supplement* is a major resource for data on the nation's social insurance and welfare programs. The majority of the statistical tables present information about the programs we administer. In 2009, our Office of Research, Evaluation and Statistics began a three-year initiative to modernize how we produce statistical tables and publications. The new process will include the development of summary data sets that can be converted to machine-readable format (e.g., comma-separated value). Consistent with our commitment to transparency, we will include the new datasets in the Data.gov inventory. As reflected in our Open Government Plan, as well as Chart 2 of this document, we plan to release the first dataset based upon this new modernized process in FY 2011. As the modernization continues, and new datasets become available, we will review and include them in the Data.gov inventory.

One of our ongoing initiatives that foster the use of our data is the Retirement Research Consortium (RRC). The Retirement Research Consortium (RRC) consists of three multidisciplinary centers housed in three separate institutions (Boston College, the University of Michigan, and the National Bureau of Economic Research). We fund the activity through cooperative agreements and awarded approximately \$7.5 million to the RRC in FY 2009, the last year of the current five-year award. We expect funding to continue at that level for each of the remaining years of the award.

The RRC has three main goals:

- Conduct research and evaluation on a wide array of topics related to Social Security and retirement policy;
- Disseminate information on Social Security and retirement issues relevant to policy makers, researchers, and the general public; and

- Train scholars and practitioners in research areas relevant to Social Security and retirement issues.

To meet these goals, the centers perform many activities. They conduct research, prepare policy briefs and working papers, hold an [annual conference](#), and provide research and training support for young scholars. A link to recent RRC research is provided below:

<http://www.ssa.gov/policy/rrc/subjects.html>

2. Other Traditional Releases from the Office of the Chief Actuary

The Trustees Report and other actuarial information is available at:

<http://www.socialsecurity.gov/OACT/pubs.html>

Other useful data are currently available from the Office of the Chief Actuary using the following link:

<http://www.socialsecurity.gov/OACT/ProgData/icp.html>

This link provides drill down capability (from 1967 forward) for information on the number of beneficiaries, average benefits, and type of beneficiaries.

Additional Future Sources and Types of Information Which May be High Value

The following chart lists potential sources and types of information that may be both available and deemed high-value in the future. This information is not currently available for public release for a number of reasons. Part of the ongoing transparency effort includes monitoring these sources and determining the appropriateness and availability of additional high-value datasets and scheduling them for release to the public.

Chart 5- Information That May Be High-Value but Not Currently Available (Planning for Release in 2013 and Continuing)

Data from public engagements and surveys and other social media

Citizen Authentication Statistics

Facility Information (subject to national security considerations)

Geographic Information System Data Presentations

Information Technology Hardware and Software Inventories

Representative Payee

Debt Management

Concluding Thoughts

This document provides the results of our complete data inventory in accordance with the Open Government Directive. However, our information and data will continue to evolve. Therefore, we will need to refresh the inventory periodically as new programs, applications, research, surveys, and public engagements change our data holdings. In some cases, we will retire and consolidate data holdings.

In addition, until the launch of Data.gov, we developed our methods and processes for gathering data, extracting data, and displaying data to support internal management, operational, and executive needs. For example, within the organization today, we do not build data gathering for public release into the project life cycle, and extraction methods have been primarily for executive decision-making or specific research. Today, where the data are available, we will reformat and present them in ways that benefit the public, while at the same time protecting their personal information.

These methods will need to evolve to incorporate transparency of the data (i.e. public release in standard formats) into the requirements for new data gathering efforts, with an ultimate goal of operating with a more agile process that considers both internal and external use of data in the future. In the future, we will specify the format and presentation of the data upfront so that preparation of data for public release will be both quicker and easier. We are exploring putting the appropriate tasks into the Planning and Analysis (P&A) and Construction phases of our systems project life cycle to identify and verify the data that would be published as well as to conduct data validation and certification. Our goal is to integrate this process with other appropriate data, testing and certification activities rather than add another stand-alone activity.

Appendix--Checklist for Data.gov Submission

NOTE: This checklist is the culmination of data review and oversight processes and is completed by the data sponsor with sign off by designated officials. Additional information for any item may be included in a supplemental page and attached to the form.

Dataset Name & Description			
*High-value {Y/N} (If Y, provide a brief explanation geared to the citizen reader of why it is high-value):			
1	<input type="checkbox"/>	The dataset is confined to public information and does not contain National Security information or other information/data protected by statute, Agency practice, legal precedent, or otherwise restricted by the SSA.	
2	<input type="checkbox"/>	The dataset complies with required privacy, confidentiality, integrity, and available controls for SSA following relevant NIST and OMB guidance.	
3	<input type="checkbox"/>	The dataset meets SSA's Information Quality Data Guidelines.	
4	<input type="checkbox"/>	The dataset does not, and should not include controls over its end use but the Federal Government cannot vouch for the data or analysis derived from these data after the data have been retrieved from Data.gov.	
5	<input type="checkbox"/>	The dataset is a product of SSA, currently available on SSA's website, and suitable for listing and downloading through one of the Data.gov catalogs. The "Raw" Data Catalog provides an instant download of machine readable, platform-independent datasets while the Tools Catalog provides hyperlinks to agency tools.	
6	<input type="checkbox"/>	The dataset is for the Data.gov Raw Data Catalog: The format of the dataset(s) is in one of the following: XML, CSV/TXT, KML/KMZ, Excel, and ESRI Shapefile or other machine readable formats (data in HTML and PDF files are not currently considered for publishing as raw data for the Data.gov "Raw" Data Catalog).	
7	<input type="checkbox"/>	Tools Catalog: If a single or multiple raw datasets are offered within a TOOL environment, the tool that offers the raw dataset(s) is one of the following: (1) Data Extraction Tool or Web Page with downloadable Datasets; (2) Feeds such as RSS, Atom or CAP; (3) Widget (tools that require logins or restrict use of raw datasets are currently not considered for publishing in the DATA.gov Tools Catalog).	
8	<input type="checkbox"/>	The data sponsor understands they are responsible for hosting data submissions. They will provide an active URL which Data.gov will only reference (i.e. no data is uploaded to DATA.gov).	
9	<input type="checkbox"/>	The data sponsor agrees to maintain the dataset and respond to all public comments.	
10	<input type="checkbox"/>	The data sponsor will submit updates to the dataset, metadata, and necessary URL (s) in a timely manner.	
11	<input type="checkbox"/>	Complete and thoroughly describe impact assessment below (Security, Privacy, FOIA, Legal):	
<p><i>*Information is high-value if it can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to a need and demand as identified through public consultation.</i></p> <p><i>** Authoritative data source is a recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent users. An authoritative data source may be the functional combination of multiple separate data sources.</i></p>			
12. Name: Data Sponsor and POC		Office	Position
Signature/Date		Phone	E-Mail
13. Name: Data Sponsor Manager		Office	Position
Signature/Date		Phone	E-Mail
14. Name: Data Quality Certifier		Office	Position
Signature/Date		Phone	E-Mail
15. Name: CIO Executive Designee and Data.gov POC		Office	Position
Signature/Date		Phone	E-Mail